DOCUMENT RESUME

ED 022 152 AL 001 413

By-Lakoff, George
COUNTERPARTS, OR THE PROBLEM OF REFERENCE IN TRANSFORMATIONAL GRAMMAR.
Pub Date 27 Jul 68
Note-13p.; Paper delivered at Linguistic Society of America Meeting, July 27, 1968, under title "Semantics, Logic, and Opacity."
EDRS Price MF-$0.25 HC-$0.60
Descriptors-*DEEP STRUCTURE, *PRONOUNS, SENTENCES, SENTENCE STRUCTURE, *STRUCTURAL ANALYSIS, *SYNTAX, *TRANSFORMATION THEORY (LANGUAGE)
Identifiers-*Counterpart theory

The author feels that although the problem of pronominalization and of reference in general is at the very heart of syntactic investigation, transformational grammarians are unable to deal adequately with these problems. He finds their theory of referential indices (Chomsky, "Aspects of the Theory of Syntax") "completely beyond repair." In this paper he presents a few examples of the failure of referential index theory and indicates the direction in which he thinks a solution will ultimately be found: counterpart theory. (DO)

COUNTERPARTS, OR THE PROBLEM OF REFERENCE IN TRANSFORMATIONAL GRAMMAR

George Lakoff

Harvard University

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED022152

Presented at the summer meeting of the Linguistic Society of America, July 27, 1968 under the misleading title, Semantics, Logic, and Opacity.

AL 001 413

The problem of pronominalization and of reference in general is at the very heart of syntactic investigation. There is hardly an area of grammar that does not impinge in some way on the problem of reference. Deletion rules involve a specification of reference. Movement rules involve constraints on reference. Deep structure constraints involve reference. Output conditions involve reference. It is impossible to investigate any but the simplest sentences in any detail without coming up against the problem of reference in one form or another.

The best that transformational grammarians have been able to come up with so far in their attempt to deal with the problems of reference is the theory of referential indices, as proposed by Chomsky in Aspects and refined recently by Postal and McCawley. The classic cases where referential indices have been used are the following:

(1) a. Pronominalization -- A pronoun that refers to its antecedent is assumed to to have the same referential index as its antecedent. Though versions of the theory vary, it is generally agreed that reflexive pronouns are to be handled in this way, and that the rule of reflexivization requires identity of referential indices.

b. Equi-NP-deletion -- In cases like I enjoyed robbing that bank, it is usually said that the deep structure subjects of enjoy and rob have the same reference, and it is assumed that they are marked with the same referential index as the subject of enjoy.

c. Relative clause formation -- It is assumed that this rule will apply when the head of the relative clause is marked with the same reference index as the noun phrase underlying the relative pronoun. In The man who left was tall, man and who would have the same referential index.

According to the index theory, if two NPs have the same referential indices, then they have the same 'intended reference'. Presumably, referential indices would be

present in semantic representations as well, so that any difficulties with the theory of referential indices will also pose difficulties for semantic theory.

The theory of referential indices is woefully inadequate, and I think it is completely beyond repair. A detailed account of the reasons for this would fill more than one course of lectures, and is certainly beyond the scope of this talk. I would like, nevertheless, to present a few of the more striking examples of the failure of this theory, and to indicate the direction in which I think a solution will ultimately be found.

Consider a sentence like (2).

(2) I dreamed that I was playing the piano.

(2) has two rather interesting readings. On reading I, which we will call the participant reading, I sense that I am seated at a piano, see the keyboard in front of me, feel my fingers hitting the keys, etc. On reading II, which I will call the observer reading, I see myself, or someone who looks just like me, sitting at a piano and playing it, as if I were watching myself in a movie. One might legitimately ask whether the fact that (2) has those two readings has any systematic linguistic interest at all. It might be the case that these two readings followed from rules of usage (if such things existed), and that the two senses of (2) should not be distinguished in either the syntactic or semantic representations of the sentence. However, there are sentences where the participant-observer distinction corresponds to a syntactic distinction.

(3) I enjoyed robbing the bank.

(4) I enjoyed my robbing the bank.

In (3), we have the participant reading, in which I enjoyed the experience of taking part in the bank robbery. In (4), we have the observer reading, in which I enjoyed observing or contemplating the event of the bank robbery which I committed. These senses can be distinguished more sharply in the following sentences.

(5) As a participant, I enjoyed robbing the bank.

(6) *As an outside observer, I enjoyed robbing the bank.

(7) *As a participant, I enjoyed my robbing the bank.

(8) As an outside observer, I enjoyed my robbing the bank.

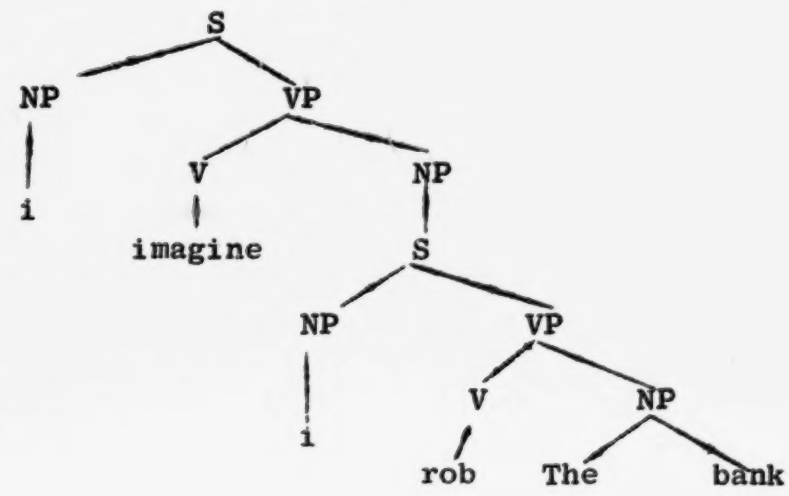
The same distinction shows up in (9) and (10).

(9) I imagined robbing the bank. /participant/

(10) I imagined myself robbing the bank. /observer/

In the theory of referential indices, one cannot even approach a description of these facts. Sentences like (9) and (10) would have the same underlying structure, such as (11).

(11)



In the derivation of (9), Equi-NP-deletion applies. In the derivation of (10), it fails to apply; then subject-raising makes the subject of rob the superficial object of imagine, and reflexivization applies, yielding myself. Equi-NP-deletion applies in the participant reading, but not in the observer reading. For the sake of Equi-NP-deletion, the subject of imagine is considered identical to the participant i, but not to the observer i. In the operation of this rule, the observer i acts like it is not co-referential with the subject of imagine. However, for the purpose of reflexivization, the subject i is considered to be identical with the subject of imagine. Thus we have the situation of (12).

(12) Equi-NP-deletion -- participant identity required

Reflexivization -- observer identity permitted

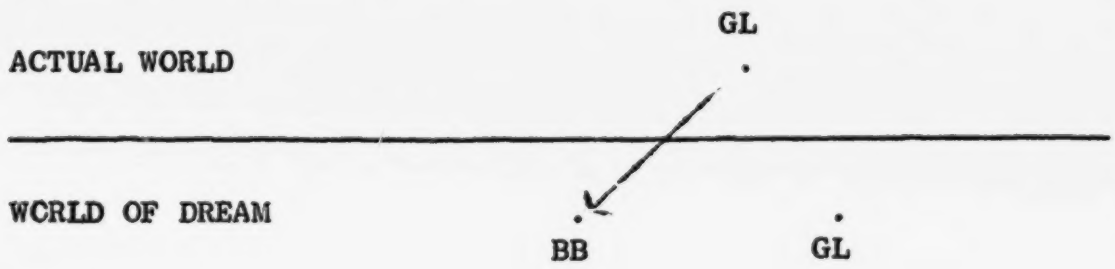
Now consider McCawley's celebrated example:

(13) I dreamed that I was Brigitte Bardot and that I kissed me.

Here we have both participant and observer identity. The I that is the subject of kiss is the participant-I and the me that is the object of kiss is the observed-I. In (13), the I that stands for Brigitte Bardot has a participant-reading, while the I that stands for ordinary me has an observed-reading. That is, (13) cannot have the reading that I dreamed that I, as an outside observer. saw Brigitte Bardot kissing someone who looked just like me. I find these facts very mysterious, and I have no idea of how to account for them. But one thing is clear. Referential indices will not do the job.

What is happening in sentences like (13) is that more than one universe of discourse or possible world is being considered. There is the actual world, in which I do the dreaming, and then there is the world of my dream. And in the world of my dream, I am split up into two people, as in (14).

(14)

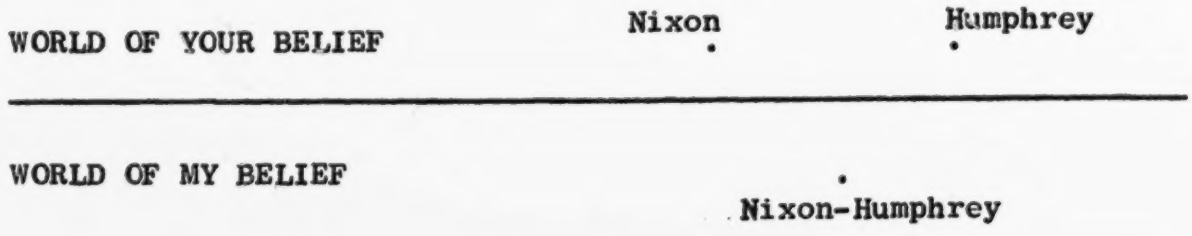


In the world of the dream I take on the consciousness of Brigitte Bardot.

Cases like (13), where one person in the actual world is split into two in some possible world, cannot be handled by the theory of referential indices. Similarly, cases where two distinct people in one possible world are collapsed into a single individual in another world also defy the referential index theory. Consider (15).

(15) You think that Nixon and Humphrey are different people and that they will campaign against each other and one of them will lose, but I think that Nixon and Humphrey are the same person and that he will win.

(16)



The complement sentence following "you think" in (15) is a statement about the world of your beliefs, where there are two distinct individuals named Nixon and Humphrey, and the plural pronoun they must be used in referring to these individuals. But the complement sentence following "I think" is a statement about the world of my beliefs where there is one individual with two names, and one must refer to that individual by the singular pronoun he. In the theory of referential indices, there is no way of making the kinds of distinction that one finds in figure (16). If Nixon and Humphrey have different referential indices in the first half of the sentence, they cannot both have the same referential index in the second half. The reason is that referential indices can only mark intended identity of reference on the part of one person -- the speaker. According to the theory, it is the speaker who intends the identity, and he cannot in the same sentence intend that two individuals be distinct, as in the first half of (15), and identical, as in the second half of (15). Thus the theory of referential indices cannot provide an adequate account of the reference in (15), on both syntactic and semantic grounds.

It should be noted that, as far as semantics is concerned, symbolic logic of the traditional sort is of no help here. The reason is that reference in logic is handled by the use of variables, which have essentially the same properties as referential indices. What one needs in order to provide semantic representations of sentences like (15) is a form of logic where one can represent possible worlds in such a way that two entities in one world can correspond to a single entity in

another world. Traditional forms of modal logic talk about possible worlds, but they usually require that the same entities appear in all possible worlds. Only recently has an approach to logic been developed which attempts to deal seriously with situations like that of figure (16). This approach is described in a very beautiful paper by David Lewis of UCLA, called "Counterpart Theory and Quantified Modal Logic," which appeared in the March 1968 issue of the Journal of Philosophy.

Of course, Lewis did not have in mind the sort of linguistic questions that we are considering. His purpose was to provide a model into which all the axioms and theorems of modal logic could be translated. Given this translation, he was able to ask whether certain axioms of modal logic whose meanings were obscure made any sense empirically. And he found that most of the axioms that modal logicians had been assuming for years were empirically ridiculous.

Lewis' system is the predicate calculus with no modal operators, but with the primitives of (17) and certain axioms governing them.

(17) Wx -- 'x is a world'

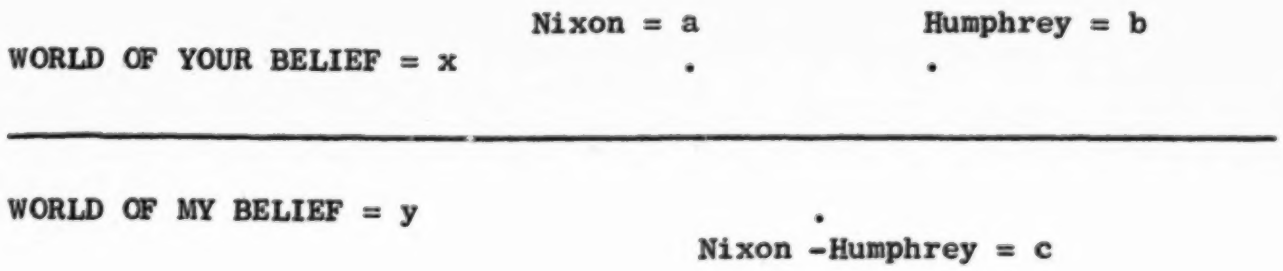
Iyx -- 'y is in x'

Ax -- 'x is actual'

Czy -- 'z is a counterpart of y'

In terms of this system we can represent the picture of (16).

(18)



(19) (i) ⌜(Ex) (Wx & Iax & Ibx)⌝ 'there is an x such that x is a world (Wx) & a is in x (Iax) & b is in x (Ibx)'

(ii) ⌜(Ey) (Wy & Icy)⌝ 'there is a y such that y is a world & c is in y'

If one considers the diagram of (16) as represented in (18), then the expressions of (19) represent the state of affairs in diagram (18). Now sentence (15) can be given a semantic representation in these terms, as in (20).

(20) (Ex)(Ey)(Ea)(Eb)(Ec) (Wx & Wy & Iax & Ibx & Icy & a is named Humphrey & b is named Nixon & (you think (Ax & (a ≠ b) & (a will run against b) & (b will run against a) & ⌜(not(a will win))or (not(b will win))⌝))
& (I think (Ay & Cac & Cbc & (c will win))))

Within a theory of semantic representation based on referential indices, such a semantic representation could not even be attempted.

There is another range of facts which cannot be described in terms of referential indices, but which can at least be approached through counterpart theory. Consider (21) - (23), which were pointed out by Leroy Baker.

(21) John wants to catch a fish and he wants to eat it.

(22) *John wants to catch a fish and he will eat it.

(23) *John wants to catch a fish and Bill wants to eat it.

If one thinks of these cases in terms of existence in possible worlds, they begin to make sense. In (21), a fish caught by John exists in the world of John's desires, which is defined by "John wants". In the second half of (21), we again have "John wants", so we are talking about the same possible world in which the fish caught by John exists, and therefore we can refer back to the fish with the pronoun "it'. In the first half of (22) we again have the world of John's desires, where there exists a fish that John caught. But in the second half of (22), we are speaking of the real world, where the fish that John wants to catch does not exist and has no counterpart. In the second half of (23) we find the same situation. We are talking about the world of Bill's desires, in which John's fish does not exist.

Now consider (24) and (25).

(24) John wants to catch a fish and he intends to eat it.

(25) *John intends to catch a fish and he wants to eat it.

Here the fish of John's desires has a counterpart in the world of his intentions, but not vice versa. Apparently, one can have intentions based on one's desires, but not desires based on one's intentions. Things that exist in one's desire-world will automatically exist in one's intention-world, but not vice versa.

Consider another set of examples.

(26) John wants to marry a girl who can beat him at chess, although he realizes that no such girl will ever exist.

(27) *John intends to marry a girl who can beat him at chess, although he realizes that no such girl will ever exist.

Here we find that entities that do not exist in the world defined by "realize" automatically do not exist in the world defined by "intend." But this is not true for "want." Thus, "realize" is related to "intend" in a way that it is not related to "want."

The same relationship holds not only for existence, but also for identity and properties.

(28) Oedipus realized that Jocasta was his mother, but he wanted to marry Jocasta without marrying his mother, ..so that they could nag each other.

(29) *Oedipus realized that Jocasta was his mother, but he intended to marry Jocasta without marrying his mother so that they could nag each other.

Here identity in the world defined by "realize" automatically entails identity in the world defined by "intend"; but this is not true for "want".

(30) John realized that he would always be rich, but he wanted to be a beggar.

(31) *John realized that he would always be rich, but he intended to be a beggar.

Here we see that if someone has a property in the world defined by "realize", he

will automatically have it in the world defined by "intend"; but not so for "want."

How can we account for facts like these? Using counterpart theory, we can define a relation, R, that holds between possible worlds, as in (32).

(32) $R(x,y) \equiv \lceil Wx \cdot Wy \cdot ((a)(Iax \supset \lceil (Eb)(Iby \cdot Cba \cdot \lceil \sim (Ed)(Idy \cdot (d \neq b) \cdot Cda) \rfloor) \rfloor) \cdot ((a)(c) \lceil (Iax \cdot Icx \cdot (c \neq a)) \supset ((b)(d) ((Iby \cdot Idy \cdot Cba \cdot Cdc) \supset (b \neq d))) \rfloor) \cdot ((a)(b) ((Iax \cdot Iby \cdot Cba \cdot Fa) \supset Fb)) \rfloor$

Briefly, (32) says the following:

(33) If world x bears the relation, R, to world y, then

i. every entity in x has a unique counterpart in y

ii. if two entities are nonidentical in x, their counterparts in y are nonidentical

iii. if an entity has a property in world x, its counterpart has that property in world y.

Thus sentences (24) and (25) show that R(want, intend), but not R(intend, want). And sentences (26) - (31) show that R(realize, intend). Sentences (34) and (35) show that R(intend, realize) does not hold.

(34) John realizes that he will find a witch and he intends to marry her.

R(realize, intend)

(35) *John intends to find a witch and he realizes that he will marry her.

$\sim$R(intend, realize)

Even given the definition of (32), it is by no means clear what it means to say that a verb taking a complement defines a possible world. Such a concept seems to be necessary, but it needs to be explicated. For example, it seems to be the case that the subjects of the verbs need to be identical for R to obtain; but it is not clear whether they can be counterparts in different worlds rather than having to be

co-referential and in the same world. But even with such uncertainties, it is possible to investigate the semantic classes of verbs by seeing what generalizations there are that involve the relation, R. I believe that this will be a very rewarding field of semantic investigation.

I assume that an adequate account of pronominalization will involve the notion counterpart in place of co-referentiality. However, I have no clear idea at present how to integrate such a notion into syntax. Moreover, even with as powerful a notion as counterpart, the participant-observer distinction of sentences (2) - (10) could not be handled. One might invent new primitives 'participant-counterpart' and 'observer-counterpart', but I think that this would just be giving the problem a name rather than solving it.

Let me conclude with a mystery about relative clauses which I hope that counterpart theory will eventually enable us to understand, but which is at present unexplained. Lauri Kartunnen has pointed out that the possibilities for relative clause formation seem to depend on the possibilities for pronominalization. Consider the following:

(36) John will catch a fish and he will eat it.

(37) John will eat the fish he catches.

(38) *John wants to catch a fish and he will eat it.

(39) *John will eat the fish he wants to catch.

This seems to gibe with Postal's theory that relative clauses are derived from conjoined sentences containing indefinite noun phrases. Postal calls this process 'repetition binding' and attempts to motivate it by sentences like the following.

(40) John kissed a nurse and she smiled.

(41) John kissed a nurse and the nurse that John kissed smiled.

But now consider sentences like (42) - (44).

(42) John wants to find a witch and he intends to marry her.

(43) *John intends to marry the witch he wants to find.

(44) *John wants to find a witch and he intends to marry the witch he wants to find.

These sentences seem to indicate that relative clause formation involves even more mysteries than pronominalization.

It should be pointed out, in conclusion, that adopting counterpart theory as a device for semantic representation may force one into adopting a philosophical position that most modern philosophers would shrink away from in horror. Consider (45).

(45) I dreamed that I found a round square and that I sold it for a million dollars.

(45) is a perfectly grammatical sentence of English. In order to account for the pronoun it in (45), one would have to say that the world of my dream is a possible world, and that a round square exists in that world. But most philosophers would maintain that round squares are logically impossible entities and therefore cannot exist in any possible world. Thus, adopting counterpart theory for the purpose of semantic theory and claiming that counterpart theory used in this way reconstructs the notion 'possible world' leads one to the position that possible worlds can contain contradictions. I think this is a perfectly reasonable position. And anyone who doesn't think that a possible world as ordinarily conceived can contain contradictions should read Catch-22, or the daily newspaper.

REFERENCES

Chomsky, Noam. Aspects of the Theory of Syntax, MIT Press, 1965.

Heller, Joseph. Catch-22. Dell paperback, 1962.

Kartunnen, Lauri. "What Do Referential Indices Refer To?" RAND Report, May, 1968.

Lewis, David. "Counterpart Theory and Quantified Modal Logic," Journal Of Philosophy, March, 1968.

McCawley, James D. "Where Do NPs Come From?" To appear in Rosenbaum and Jacobs, Readings in English Transformational Grammar, 1968.

Postal, Paul. "Crazy Notes on Restrictive Relatives." Unpublished, Spring 1967.

Postal, Paul. "Notes on Repetition Binding." Unpublished, 1968.